HSR Health Services Research

DOI: 10.1111/1475-6773.12089
METHODS ARTICLE

Using the Bayesian Improved Surname Geocoding Method (BISG) to Create a Working Classification of Race and Ethnicity in a Diverse Managed Care Population: A Validation Study

Dzifa Adjaye-Gbewonyo, Robert A. Bednarczyk, Robert L. Davis, and Saad B. Omer

Objective. To validate classification of race/ethnicity based on the Bayesian Improved Surname Geocoding method (BISG) and assess variations in validity by gender and age.

Data Sources/Study Setting. Secondary data on members of Kaiser Permanente Georgia, an integrated managed care organization, through 2010.

Study Design. For 191,494 members with self-reported race/ethnicity, probabilities for belonging to each of six race/ethnicity categories predicted from the BISG algorithm were used to assign individuals to a race/ethnicity category over a range of cutoffs greater than a probability of 0.50. Overall as well as gender- and age-stratified sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. Receiver operating characteristic (ROC) curves were generated and used to identify optimal cutoffs for race/ethnicity assignment.

Principal Findings. The overall cutoffs for assignment that optimized sensitivity and specificity ranged from 0.50 to 0.57 for the four main racial/ethnic categories (White, Black, Asian/Pacific Islander, Hispanic). Corresponding sensitivity, specificity, PPV, and NPV ranged from 64.4 to 81.4 percent, 80.8 to 99.7 percent, 75.0 to 91.6 percent, and 79.4 to 98.0 percent, respectively. Accuracy of assignment was better among males and individuals of 65 years or older.

Conclusions. BISG may be useful for classifying race/ethnicity of health plan members when needed for health care studies.

Key Words. Race/ethnicity, imputation and indirect estimation, geocoding, surname analysis, health plans

Racial and ethnic differences and disparities in health and health care are issues of growing concern that need to be properly identified and classified to be adequately studied and addressed. Yet health plans often do not readily have race/ethnicity information at their disposal for a large proportion of their members (High-Value Health Care Project 2010; Weissman and Hasnain-Wynia 2011). This may be due to limitations, such as infrequent contact with members, added costs of developing the systems to collect such information, perceptions of regulatory prohibitions against collecting race/ethnicity, and discomfort or fear among members in providing it (America's Health Insurance Plans and Robert Wood Johnson Foundation's 2004; Institute of Medicine 2009; Escarce et al. 2011; Weissman and Hasnain-Wynia 2011; Gazmararian et al. 2012).

Although many health plans are beginning to collect race/ethnicity data more systematically, it is likely to take a substantial amount of time before this information is complete (Escarce et al. 2011). While self-report is considered the gold standard (Institute of Medicine 2009; Gazmararian et al. 2012), several techniques have been developed to indirectly estimate race/ethnicity where it is unavailable, which would allow for monitoring and evaluation of its role in health care and health research in the meantime (High-Value Health Care Project 2010). This approach has been encouraged by the United States (US) government's Agency for Healthcare Research and Quality and the Institute of Medicine (Institute of Medicine 2009). One such recent method is the Bayesian Improved Surname Geocoding method (BISG) developed by Rand Corporation, which has been demonstrated to improve upon previous techniques (Elliott et al. 2009). The algorithm, described in detail by Elliott et al. (2008, 2009), utilizes a Bayesian approach to combine racial/ethnic data from last names and geographic units.

While the use of indirect race/ethnicity estimations can enhance research efforts, an assessment of the quality of these data in a variety of contexts is needed before implementation within each setting. Some validation studies comparing the BISG algorithm estimates to self-reported race/ethnicity have been performed by its developers (Elliott et al. 2008, 2009; Elliott

---

Address correspondence to Dzifa Adjaye-Gbewonyo, MPH, Center for Health Research-Southeast, Kaiser Permanente, 11 Piedmont Center, 3495 Piedmont Road NE, Suite 110, Atlanta, GA 30305; e-mail: Dzifa.E.Adjaye@kp.org. Robert A. Bednarczyk, Ph.D., Robert L. Davis, M.D., M.P.H., and Saad B. Omer, M.B.B.S., M.P.H., Ph.D., are with the Center for Health Research-Southeast, Kaiser Permanente, Atlanta, GA. Robert A. Bednarczyk, Ph.D., and Saad B. Omer, M.B.B.S., M.P.H., Ph.D., are with the Rollins School of Public Health, Emory University, Atlanta, GA.

2009; High-Value Health Care Project 2010) as well as by health plans beginning to utilize it. However, these studies have included limited additional stratification by pertinent personal characteristics; in particular, analysis of the BISG algorithm's performance by age is lacking. Moreover, while the intended use of the BISG algorithm is to directly apply the predicted probabilities and not to classify individuals into a single race/ethnicity, it is important to understand how the algorithm performs for individual assignment of race/ethnicity in instances where the use of probability values is not feasible. For example, when attempting to identify individuals of a certain race/ethnicity for participation in a targeted program or research study, it will be necessary to decide an appropriate probability level that would warrant contacting individuals for recruitment. Yet analyses to determine optimal cutoffs on which to base such determinations have not been reported. In this study, we therefore assessed the validity of BISG algorithm estimations for the "off-label" use of assigning race/ethnicity within a regional managed care organization (MCO) population and examined how well the algorithm performed by gender and age.

METHODS

Study Population

The study population consisted of members of Kaiser Permanente's Georgia region (KPGA)—one of eight regions of the Kaiser Permanente Medical Care Program, a federally qualified, prepaid group health maintenance organization. The study population was limited to those who had self-reported race/ethnicity information available in Kaiser Permanente's electronic medical record (EMR) system databases and had a geocodable address allowing race/ethnicity to be calculated using BISG. Self-reported race/ethnicity is primarily collected during medical encounters and is available on roughly 80 percent of the current KPGA membership. The time frame used for the study population was through December 31, 2010.

Data were sourced from Kaiser Permanente's Geographically Enriched Member Sociodemographics datamart, which contains any available race/ethnicity data reported for all medical record numbers in the EMR system in both a raw and standardized form as well as imputed race/ethnicity information calculated using BISG. The BISG algorithm results in individual-level probability distributions indicating the likelihood of belonging to each of six mutually exclusive race/ethnicity categories. Race/ethnicity data for each members' census block group (based on Esri's Updated Census Demographics dataset, 2010 [Esri 2012]) and members' last names (based on the 2000 Census Bureau surname list [U.S. Census Bureau 2012]) were used for the BISG assignment process.

Analysis

Members were assigned to one of the six races/ethnicities (White, Black, Asian or Pacific Islander [API], Hispanic, American Indian or Alaska Native [AIAN], Multi-racial) using specified cutoff levels for the individual-level race/ethnicity probabilities imputed by the BISG algorithm. Meeting a given cutoff criterion was defined as having a predicted probability for a race/ethnicity greater than the cutoff value, while probabilities less than or equal to the cutoff value failed to meet the criterion. The lowest cutoff was specified as 0.50 (or 50 percent) to eliminate any possibility of an individual meeting the cutoff for more than one category, and this cutoff was varied in increments of 0.01 (1 percent) to a maximum cutoff of 1 (100 percent). Assigned race/ethnicity was validated against reported race/ethnicity for the study population by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The proportion of the study population meeting each cutoff level, as well as the overall accuracy of race/ethnicity assignment—defined as the percentage of individuals at each cutoff level assigned to the correct reported race/ethnicity, was also determined.

Sensitivity and specificity results were used to generate a receiver operating characteristic (ROC) curve for each race/ethnicity, and an optimum cutoff level for the assignment of each race/ethnicity was determined by selecting the cutoff point on the ROC curve with the minimum Euclidean distance from the point of perfect prediction (100 percent sensitivity and specificity). The formula for this calculation is derived from the Pythagorean theorem, where $\sqrt{(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2}$ is the Euclidean distance to the point of perfect prediction from the point corresponding to the cutoff level in question (Akobeng 2007; Gönen 2007).

The above processes were repeated with the population stratified by gender to assess the impact of gender on the algorithm's performance. To examine variations in validity by age, each member's age in years, as of December 31, 2010, was calculated and was categorized as follows: 0 to less than 18 years, 18 to less than 50 years, 50 to less than 65 years, and 65 years and older. These broad categories were designed to focus primarily on extreme age groups (children and the elderly), with the remaining adults divided into two categories. Children under 5 years were also analyzed separately as a subset of the less than 18 age group. In the case of minors, race may have been reported by the parent or guardian rather than through direct self-report. All analyses were conducted in SAS 9.2 (The SAS Institute, Cary, NC, USA).

RESULTS

Population-Level Characteristics and Algorithm Outcomes

The study population consisted of a total of 191,494 individuals. Blacks represented the largest proportion, followed closely by Whites. A greater proportion was female, and adults 65 years and older represented the smallest age group (Table 1).

In this population, individual predicted probabilities from the BISG algorithm for each race/ethnicity had a mean of 0.475 for White, 0.393 for Black, 0.065 for Hispanic, 0.050 for API, 0.002 for AIAN, and 0.014 for Multi-racial. Individuals' maximum predicted probabilities ranged from 0.245 to 1.000 with a median of 0.862.

Validity of Individual-Level Race/Ethnicity Predictions

Most of the study population (96.1 percent) met the starting cutoff of 0.50 for the assignment of race/ethnicity, decreasing to 0 percent at a maximum cutoff of 1. Therefore, the proportion of the total population that was accurately assigned decreased from 76.4 to 0 percent over the cutoff range. However, the proportion of the population meeting each cutoff whose race/ethnicity was accurately assigned increased from 79.5 percent at 0.50 to 98.2 percent at 0.99.

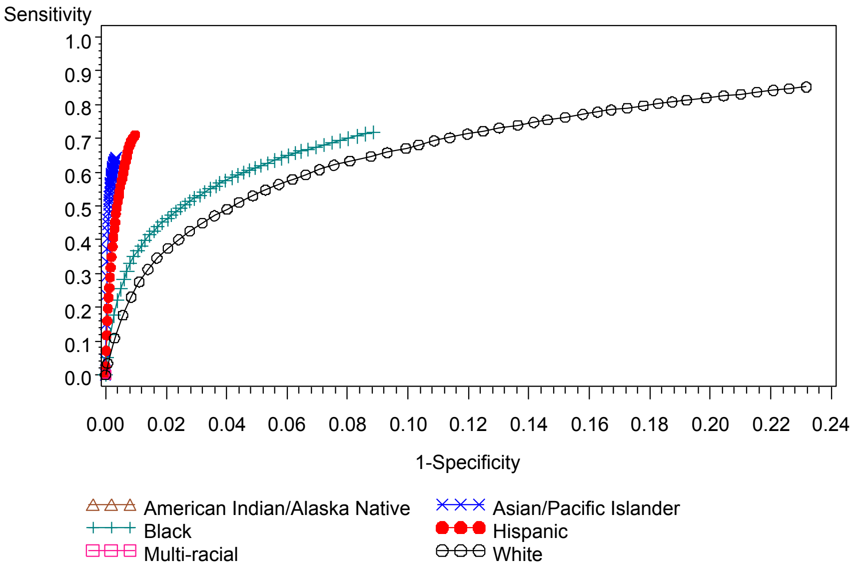
Sensitivity at a cutoff of 0.50 for White, Black, Hispanic, API, AIAN, and Multi-racial, respectively, was 85.2, 71.8, 71.0, 64.4, 0.0, and 0.3 percent, with each decreasing to a value of 0 percent at a cutoff of 1. Specificity increased from 76.8, 91.1, 99.0, 99.6, 100.0, and 100.0 percent, respectively, at a cutoff of 0.50 to 100 percent at a cutoff of 1. The ROC curves (Figure 1) illustrate the changes in sensitivity and specificity and the relationship between the two for each race/ethnicity.

From a cutoff of 0.50 to 0.99, PPV increased from 72.2 to 97.2 percent for White; 87.2 to 98.8 percent for Black; 83.4 to 97.5 percent for Hispanic; and 91.6 to 99.4 percent for API. PPV was 0 percent for AIAN through a

Table 1: Racial/Ethnic Distribution of the Study Population as a Whole and Stratified by Gender and Age

	Total	*Gender*		*Age*			
	(N = 191,494) (%)	*Male (N = 82,320, 43.0%)*	*Female (N = 109,174, 57.0%)*	*0 to < 18 years (N = 38,664, 20.2%)*	*18 to < 50 years (N = 92,328, 48.2%)*	*50 to < 65 years (N = 43,239, 22.6%)*	*65+ years (N = 17,263, 9.0%)*
Race/ethnicity							
White	41.4	45.0	38.7	36.8	37.3	47.3	59.1
Black	45.7	41.8	48.7	48.5	48.1	42.9	33.8
Hispanic	6.5	6.6	6.4	6.7	8.1	4.2	3.4
Asian/Pacific Islander	5.6	5.8	5.5	6.9	5.9	4.8	3.0
American Indian/Alaska Native	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Multi-racial	0.6	0.7	0.6	0.9	0.6	0.6	0.5

Figure 1: Receiver Operating Characteristic Curves for the BISG Algorithm by Race/Ethnicity



cutoff of 0.53 and was otherwise undefined. For Multi-racial, PPV fluctuated between 0 and 13.8 percent through a cutoff of 0.73 and was thereafter undefined. Over the range of cutoffs from 0.50 to 1, NPV decreased from 88.0 to 58.6 percent for White; 79.3 to 54.3 percent for Black; 98.0 to 93.5 percent for Hispanic; and 97.9 to 94.4 percent for API. NPV decreased only slightly for Multi-racial, remaining around 99.4 percent; and for AIAN NPV was constant at 99.9 percent over all cutoff levels.

Optimal Cutoff Selection

Within the range of cutoffs examined, the prediction of race/ethnicity was optimized to most closely approximate perfect prediction based on a combination of sensitivity and specificity at 0.57 for the assignment of White. In other words, valid prediction of White race would be optimized using 0.57 as the criterion for assignment and assigning any individual whose predicted probability from BISG for being White was greater than 0.57. Other optimized values were any cutoff starting from 0.54 for the assignment of AIAN—as all cutoffs from that value onwards were equivalent with specificity having reached 100 percent while sensitivity was consistently 0 percent—and the minimum cutoff of 0.50 for the assignment of all other categories (Black, Hispanic, API, and Multi-racial). Corresponding measures of sensitivity, specificity, PPV, and NPV at these optimal cutoffs for each race/ethnicity are presented in Table 2.

Stratification by Gender

The racial/ethnic distribution of the study population differed by gender and consisted of more White males than females and more Black females than males (Table 1). Other racial/ethnic groups were similarly distributed across genders.

Males and females, 96.1 and 96.0 percent, respectively, had a predicted probability meeting the minimum cutoff of 0.50, which decreased with increasing cutoff. Thus, the proportion of the total male and female population who was accurately assigned decreased from 77.4 and 75.6 percent, respectively, to 0 percent over the range of cutoffs; but of those who met each cutoff level, the proportion whose race/ethnicity was accurately assigned ranged from 80.6 percent at 0.50 to 98.8 percent at 0.99 for males and from 78.7 to 97.8 percent for females (Figure 2).

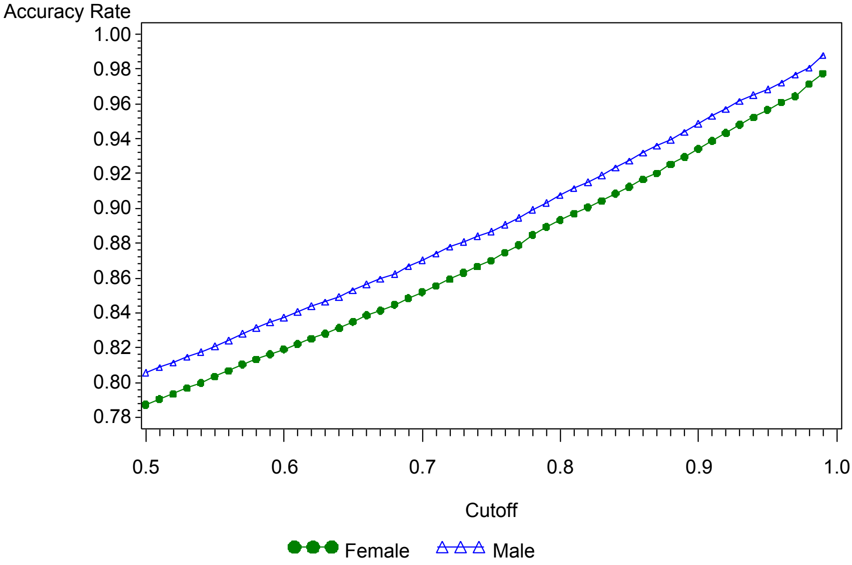
The cutoff that optimized the prediction of each race/ethnicity for males and females, respectively, was 0.57 and 0.56 for White; any cutoff starting from 0.52 and from 0.54 for AIAN; and 0.50 for all other races/ethnicities. Gender-stratified sensitivities, specificities, PPVs, and NPVs for each race/ethnicity over the range of cutoffs and at the optimal cutoff are contained in the Supplement (Table S1).

Table 2: Validity of Predictions by Race/Ethnicity at the Identified Optimal Cutoffs

Race/Ethnicity	*Optimal Cutoff*	*Sensitivity (%)*	*Specificity (%)*	*PPV (%)*	*NPV (%)*
White	0.57	81.4	80.8	74.9	86.0
Black	0.50	71.8	91.1	87.2	79.3
Hispanic	0.50	71.0	99.0	83.4	98.0
Asian/Pacific Islander	0.50	64.4	99.6	91.6	97.9
American Indian/ Alaska Native	0.54 or higher	0	100	und	99.9
Multi-racial	0.50	0.3	100.0	13.8	99.4

NPV, negative predictive value; PPV, positive predictive value; und, undefined.

Figure 2: Proportion of the Study Population with Accurate Predictions by Sex as a Function of the Cutoff Used for Race/Ethnicity Assignment from BISG Probabilities

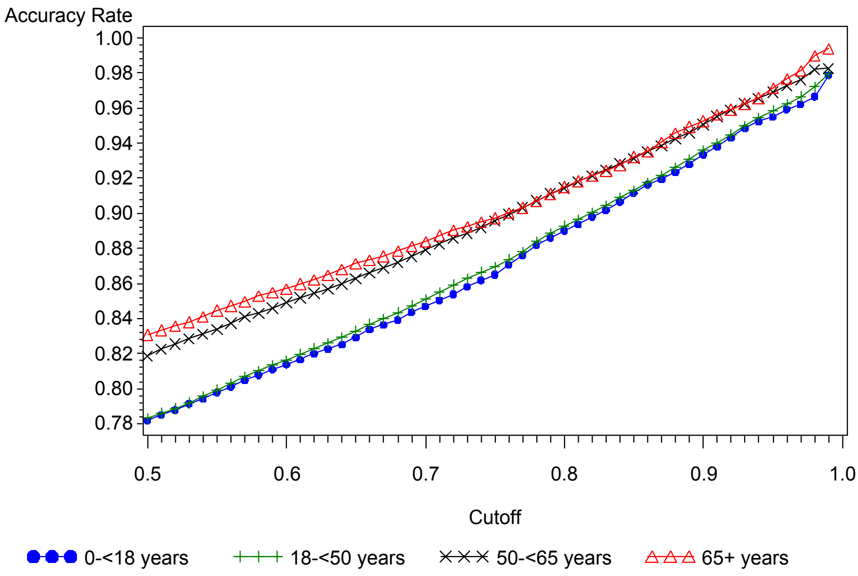


Stratification by Age

The proportion of non-Whites in the study population was greater in younger age groups (Table 1). In particular, the proportion of Whites increased with members' age while the pattern was reversed in Blacks and APIs. Hispanics and Multi-racial individuals had a similar decrease with age, although not following a consistent pattern, while AIANs were relatively stable.

Of people aged 0 to 18, 18 to 50, 50 to 65, and 65 or older, respectively, 96.0, 95.7, 96.6, and 96.9 percent had a predicted probability meeting the initial 0.50 cutoff. As the cutoff increased, these proportions decreased, and the proportion of the total population accurately assigned by age group decreased from 75.1, 74.9, 79.1, and 80.5 percent, respectively. However, the age-stratified rates of accurate prediction of race/ethnicity among those meeting each cutoff from 0.50 to 0.99 ranged from 78.2 to 97.9 percent in 0 to 18 year olds, 78.3 to 98.0 percent in 18 to 50 year olds, 81.9 to 98.3 percent in 50 to 65 year olds, and 83.1 to 99.4 percent in adults 65 years and older (Figure 3). Optimal cutoffs by age for the assignment of race/ethnicity as well as age-stratified

Figure 3: Proportion of the Study Population with Accurate Predictions by Age Group as a Function of the Cutoff Used for Race/Ethnicity Assignment from BISG Probabilities



sensitivities, specificities, PPVs, and NPVs are presented in the Supplement (Table S2).

For the subset of children under 5 years of age (5.6 percent of the study population), 95.3 percent met the initial cutoff of 0.50, and the accuracy rate for predictions over the cutoff range was from 77.5 to 96.9 percent. Optimal cutoff values were similar to those found in the 0–18 year age group.

Discussion

We found that the Bayesian Improved Surname Geocoding algorithm performed well in this population for the assignment of individual race/ethnicity based on predicted probabilities. This conclusion is based in particular on the high overall accuracy rates for predictions among those meeting cutoffs starting from 0.50, which were around 80 percent or higher. Moreover, despite the fact that the number of individuals with high enough predicted probabilities drops as the cutoff is raised, 89–96 percent of the initial population was still retained at the overall cutoffs levels determined to be optimal for this population (0.50 to 0.57).

The algorithm was most sensitive for White, but extremely insensitive in identifying AIAN or Multi-racial individuals, with sensitivities among all categories converging at higher cutoffs. As a result of their low sensitivity, the algorithm had perfect specificity for AIAN and Multi-racial; however, specificity was likewise high for API and Hispanic (99 percent or above) and was above 90 percent for Black.

Based on the placement of the ROC curves, algorithm performance appears best for Hispanic and API—due to little change in specificity with sensitivity and thus closer proximity to the top left point designating perfect prediction (1, 0)—followed next by Black and White. Although these results may differ slightly from findings on the BISG algorithm published by Elliott et al. (Elliott et al. 2009), because of the differences in the insured populations as well as different validation methods, they are nonetheless comparable to their findings of higher correlation for Hispanic and Asian. The poor performance for AIAN and Multi-racial is likewise consistent with their results and indicative of the low utility of surname analysis and geocoded residence in distinguishing these populations (Elliott et al. 2009).

The study population's differing racial/ethnic distribution by gender either suggests differential reporting of race/ethnicity by gender or differential enrollment in the health plan due to differences between men and women of various races/ethnicities in employment patterns and insurance coverage. Stratification by gender also revealed that algorithm performance was generally slightly better in males. Gender-stratified ROC curves (not presented here) were shifted higher in males compared to females in the four major race/ethnicity categories, although they were unstable in AIAN and Multi-racial. This difference is presumably a result of the common practice of women adopting their husbands' last name at marriage, and given that this study population was limited to one region of the United States, these differences may also suggest that women in the Southeast are more likely to change their last names than women in the nation as a whole. A slight gender difference in validity is supported in the literature on BISG (Elliott et al. 2009) as well as other applications of indirect estimation of ethnicity involving surname analysis (Howard et al. 1983; Hazuda et al. 1986; Fiscella and Fremont 2006; Quan et al. 2006; Wei et al. 2006; Shah et al. 2010; Wong, Palaniappan, and Lauderdale 2010; Lakha, Gorman, and Mateos 2011). In terms of sensitivity, the largest gender gap was between Hispanics followed by Asians, most likely a result of the greater contribution of surname analysis for these groups compared to others. While the algorithm was consistently more specific in males, surprisingly, among Blacks only, sensitivity and PPV were higher in women. Although Elliott et al. utilize a different measure of performance, they also find greater gender differences in Hispanic and Asian and a reverse difference for Black (Elliott et al. 2009).

The trends in the population's racial/ethnic distribution with age suggest differences in enrollment between elderly and young Whites and non-Whites; namely, elderly non-Whites appear less likely to be covered in this health plan. It may also indicate some increasing diversity of the insured population over time or differential utilization and reporting of race/ethnicity by age.

As a whole, the algorithm appeared to perform better in older age groups, possibly suggesting that older individuals may live in less diverse communities and carry last names that are more suggestive of their race/ethnicity. Furthermore, this pattern may be stronger in the Southeastern U.S. population used for this analysis than in the national population. Similar results demonstrating improved performance with age among Asians and Hispanics using name lists have also been found in some studies (Hazuda et al. 1986; Shah et al. 2010; Wong, Palaniappan, and Lauderdale 2010). However, by race/ethnicity, the algorithm was less sensitive with age for Whites in this study, perhaps reflecting decreased residential mobility of older Whites who thus remain in increasingly integrated neighborhoods (Sandberg et al. 2009). Conversely, sensitivity increased with age for Blacks, suggesting that older individuals may remain in more segregated, predominantly Black neighborhoods while younger individuals are more likely to move into more diverse areas. A similar increasing pattern was seen for Multi-racial—despite very low sensitivity—and somewhat for API, while patterns were inconsistent for Hispanic and AIAN.

The importance of considering age has been illustrated previously with a non-Bayesian algorithm combining surname analysis and geocoding that found better performance when utilizing age-stratified race/ethnicity information on geographic units rather than summary information; this recognizes the dynamic nature of neighborhoods over time and age differentials in mobility (Sandberg et al. 2009).

It is important to note that the validity results and cutoffs presented here are not meant to be generalizable as this analysis was based on a regional and not a national sample. However, the methodology utilized here can be implemented in other regions to obtain results relevant to different locations. Algorithm performance will be highly dependent on the underlying source population and the relative prevalence of different racial/ethnic groups as well as the level of racial/ethnic residential heterogeneity within the population due to its partial basis on geocoding. Thus, the algorithm may be insensitive to less prevalent races/ethnicities in populations with a predominant racial/ethnic group while increased residential segregation would allow for individual probability predictions high enough to satisfy a given cutoff (Fiscella and Fremont 2006). Given the greater racial/ethnic differences between regions of the United States, one would therefore expect better performance in a national sample than the regional sample presented in this analysis. Furthermore, the extremely low prevalence of AIAN and Multi-racial in this population may partially account for the poor sensitivity observed. The Supplement provides a quantitative example demonstrating how differences in the source population structure affect performance.

Moreover, there are alternative approaches and methods for the selection of an appropriate cutoff level for classification; therefore, how one is selected and how a race variable is defined will depend on the research question and the relative importance of false negatives and false positives.

The predicted probabilities from BISG can be used directly in regression and other analyses without the need for individual classification (Elliott et al. 2008, 2009). Utilizing the probabilities to categorize individuals can lead to a loss of information by effectively equating individuals who may have very different probabilities and can also result in reduced accuracy and efficiency (Elliott et al. 2008, 2009). However, the ability to make individual categorizations of race/ethnicity can be especially helpful for targeting communication campaigns and for improving recruitment of specific races/ethnicities for studies and other programs. Thus, an assessment of how best to make such classifications is important for this purpose.

Ultimately, the goal for health plans is to increase efforts at collecting and obtaining reported race. However, in the interim, the use of indirect methods such as BISG may be beneficial in health care and health research. With overall accuracy rates near or above 80 percent and moderate to high sensitivity (64–81 percent) as well as high specificity and predictive values (75–100 percent) at optimal cutoff levels, the reasonably high validity of the algorithm for major racial/ethnic groups within this population—even when a simple majority was the basis for individual prediction—is indicative of the fact that the algorithm may indeed be reliable for individual-level assignment within populations of a similar racial/ethnic structure.

In conclusion, the results of this study demonstrate suitable performance of BISG overall for the classification of race/ethnicity within a diverse MCO population as well as greater variation in performance by age than by gender, as seen in the differences in optimal cutoffs selected. This suggests that when applying the algorithm, it may be crucial to take the age of the population into account.

Acknowledgments

Joint Acknowledgment/Disclosure Statement: The authors thank Kaiser Permanente's Utility for Care Data Analysis (UCDA) for providing documentation and resources on the GEMS datamart used for this analysis. The authors also acknowledge that they have no conflicts of interest.

Disclosures: None.

Disclaimers: None.

References

Akobeng, A. K. 2007. "Understanding Diagnostic Tests 3: Receiver Operating Characteristic Curves." *Acta Paediatrica* 96 (5): 644–7.

America's Health Insurance Plans and Robert Wood Johnson Foundation. 2004. "Collection of Racial and Ethnic Data by Health Plans to Address Disparities: Final Summary Report" [accessed on December 7, 2011]. Available at http://www.rwjf.org/files/research/080504AHIPFinalSummary.pdf

Elliott, M. N. 2009. "Use of Indirect Measures of Race/Ethnicity to Target Disparities" [accessed on October 18, 2011]. Available at http://www.iom.edu/~/media/Files/Activity%20Files/Quality/QualityDisparities/ElliotSlides.pdf

Elliott, M. N., A. Fremont, P. A. Morrison, P. Pantoja, and N. Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43 (5 Pt 1): 1722–36.

Elliott, M., P. Morrison, A. Fremont, D. McCaffrey, P. Pantoja, and N. Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology* 9 (2): 69–83.

Escarce, J. J., R. Carreon, G. Veselovskiy, and E. H. Lawson. 2011. "Collection of Race and Ethnicity Data by Health Plans has Grown Substantially, but Opportunities Remain to Expand Efforts." *Health Affairs (Millwood)* 30 (10): 1984–91.

Esri. 2012. "Demographic Data including U.S. Population Data" [accessed on September 10, 2012]. Available at http://www.esri.com/data/esri_data/demographic-overview/demographic

Fiscella, K., and A. M. Fremont. 2006. "Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity." *Health Services Research* 41 (4 Pt 1): 1482–500.

Gazmararian, J., R. Carreon, N. Olson, and B. Lardy. 2012. "Exploring Health Plan Perspectives in Collecting and Using Data on Race, Ethnicity, and Language." *American Journal of Managed Care* 18 (7): e254–61.

Gönen, M. 2007. *Analyzing Receiver Operating Characteristic Curves with SAS.* Cary, NC: SAS Institute.

Hazuda, H. P., P. J. Comeaux, M. P. Stern, S. M. Haffner, C. W. Eifler, and M. Rosenthal. 1986. "A Comparison of Three Indicators for Identifying Mexican Americans in Epidemiologic Research. Methodological Findings from the San Antonio Heart Study." *American Journal of Epidemiology* 123 (1): 96–112.

High-Value Health Care Project. 2010. "Indirect Estimation of Race and Ethnicity: An Interim Strategy to Measure Population Level Health Care Disparities" [accessed on October 13, 2011]. Available at http://www.healthqualityalliance.org/userfiles/RAND%20issue%20brief%20031810.pdf

Howard, C. A., J. M. Samet, R. W. Buechley, S. D. Schrag, and C. R. Key. 1983. "Survey Research in New Mexico Hispanics: Some Methodological Issues." *American Journal of Epidemiology* 117 (1): 27–34.

Institute of Medicine. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement.* Washington, DC: The National Academies Press.

Lakha, F., D. R. Gorman, and P. Mateos. 2011. "Name Analysis to Classify Populations by Ethnicity in Public Health: Validation of Onomap in Scotland." *Public Health* 125 (10): 688–96.

Quan, H., F. Wang, D. Schopflocher, C. Norris, P. D. Galbraith, P. Faris, M. M. Graham, M. L. Knudtson, and W. A. Ghali. 2006. "Development and Validation of a Surname List to Define Chinese Ethnicity." *Medical Care* 44 (4): 328–33.

Sandberg, T. J., A. R. Wilson, H. Rodin, N. A. Garrett, E. P. Bargman, D. Dobmeyer, and D. W. Plocher. 2009. "Improving the Imputation of Race: Evaluating the Benefits of Stratifying by Age." *Population Health Management* 12 (6): 325–31.

Shah, B. R., M. Chiu, S. Amin, M. Ramani, S. Sadry, and J. V. Tu. 2010. "Surname Lists to Identify South Asian and Chinese Ethnicity from Secondary Data in Ontario, Canada: A Validation Study." *BMC Medical Research Methodology* 10: 42.

U.S. Census Bureau. 2012. "Genealogy Data: Frequently Occurring Surnames from Census 2000" [accessed on June 15, 2013]. Available at http://www.census.gov/genealogy/www/data/2000surnames/index.html

Wei, II, B. A. Virnig, D. A. John, and R. O. Morgan. 2006. "Using a Spanish Surname Match to Improve Identification of Hispanic Women in Medicare Administrative Data." *Health Services Research* 41 (4 Pt 1): 1469–81.

Weissman, J. S., and R. Hasnain-Wynia. 2011. "Advancing Health Care Equity through Improved Data Collection." *New England Journal of Medicine* 364 (24): 2276–7.

Wong, E. C., L. P. Palaniappan, and D. S. Lauderdale. 2010. "Using Name Lists to Infer Asian Racial/Ethnic Subgroups in the Healthcare Setting." *Medical Care* 48 (6): 540–6.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1. Validity of Gender-Stratified Predictions by Race/Ethnicity at the Optimal Cutoff and over the Cutoff Range.*

Table S2. Validity of Age-Stratified Predictions by Race/Ethnicity at the Optimal Cutoff and over the Cutoff Range.*

Appendix SA1: Author Matrix.